



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/768,947	01/24/2001	William Pugh	GOOGLE-5	4264

26479 7590 03/12/2003

STRAUB & POKOTYLO
1 BETHANY ROAD, SUITE 83
BUILDING 6
HAZLET, NJ 07730

EXAMINER

NGUYEN, MERILYN P

ART UNIT

PAPER NUMBER

2171

DATE MAILED: 03/12/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/768,947	PUGH ET AL.	
	Examiner	Art Unit	
	Marilyn P Nguyen	2171	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on _____.
- 2a) This action is FINAL. 2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) Claim(s) 1-45 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) Claim(s) 1-13, 17-29, 34-38 and 40-45 is/are allowed.
- 6) Claim(s) 14, 15, 30-33, and 39 is/are rejected.
- 7) Claim(s) 16 is/are objected to.
- 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on 24 January 2001 is/are: a) accepted or b) objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- 11) The proposed drawing correction filed on _____ is: a) approved b) disapproved by the Examiner.
If approved, corrected drawings are required in reply to this Office action.
- 12) The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

- 13) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
 - a) All
 - b) Some *
 - c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.
- 14) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).
 - a) The translation of the foreign language provisional application has been received.
- 15) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

- | | |
|--|--|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) Paper No(s). _____ . |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____ . | 6) <input type="checkbox"/> Other: _____ . |

DETAILED ACTION

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraph of 35 U.S.C. § 102 in view of the AIPA and H.R. 2215 that forms the basis for the rejections under this section made in the attached Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

35 U.S.C. § 102(e), as revised by the AIPA and H.R. 2215, applies to all qualifying references, except when the reference is a U.S. patent resulting directly or indirectly from an international application filed before November 29, 2000. For such patents, the prior art date is determined under 35 U.S.C. § 102(e) as it existed prior to the amendment by the AIPA (pre-AIPA 35 U.S.C. § 102(e)).

Claims 14, 15, 30-33, and 39 are rejected under 35 U.S.C. 102(e) as being anticipated by Broder (US 6,119,124).

Regarding claims 14 and 30, Broder discloses a method for filtering search results to remove near-duplicates, the method comprising:

a) for each of a predetermined number of candidate search results, determining whether the candidate search result is a near-duplicate of another candidate search result (See col. 9, lines 1-7, and col. 12, claim 1); and

b) if it is determined that the candidate search result is a near-duplicate of another candidate search result, then rejecting the candidate search result (See col. 9, lines 4-5).

Regarding claims 15 and 31, Broder discloses wherein the act of determining whether a candidate search result is a near-duplicate of another candidate search result includes

- i) comparing a cluster identifier of the candidate search result with that of the other candidate search result (See col. 6, lines 54-63), and
- ii) if the cluster identifiers of the two candidate search results match, then concluding that the two candidate search results are near-duplicates (See col. 5, lines 13-20).

Regarding claim 32, Broder discloses a machine-readable medium having stored thereon a plurality of records (See Fig. 4), each of the records comprising:

- a) a first field for storing a document identifier (See col. 7, lines 35-36); and
- b) a plurality of lists, each of the plurality of lists containing elements of a documents identified by the document identifier stored in the first field (See col. 7, lines 35-36, “shingle value”),

wherein a hash function is used to determine which of the plurality of lists each of the elements will be contained in (See col. 6, lines 5-7).

Regarding claim 33, Broder discloses a machine-readable medium having stored thereon a plurality of records, each records comprising:

- a) a first field for storing a document identifier (See col. 7, lines 35-36); and

b) a plurality of fingerprints (See col. 6, lines 54-63), wherein each of the fingerprints is a low collision probability hash function of elements contained in a corresponding list (See col. 6, lines 61-63), and wherein the elements are elements of a document identified by the document identifier stored in the first field (See col. 7, lines 35-36, “shingle value”).

Regarding claim 39, Broder discloses a method for determining whether two documents are near-duplicates (See col. 4, line 6 et seq.), the method comprising:

- a) for each of the two documents, generating at least two fingerprints (See col. 4, lines 19-24); and
- b) determining whether or not the two documents are near-duplicate documents by
 - i) determining whether or not any one of the fingerprints of a first of the two documents matches any one of the fingerprints of a second of the two documents (See col. 5, lines 17-19), and
 - ii) if it is determined that anyone fingerprints of the first of the two documents does match any one fingerprints of the second of the two documents, then concluding that the two documents are near-duplicates (See col. 17-19).

Allowable Subject Matter

Claim 16 is objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Art Unit: 2171

Regarding claim 16, the prior art of record fails to disclose or suggest the claimed step of associating as addresses below in claim 13.

Claims 1, 11, 13, 17, 26-29, 34, 35, 38, 40, and 44 are allowed.

Regarding claims 1 and 11, the prior art of record fail to disclose or suggest the claimed steps of: preprocessing the fingerprints to identify any fingerprints that are associated with only one document and determining whether or not documents are near-duplicate documents based on fingerprints other than those identified as being associated with only one document in conjunction with the remaining, salient claim provisions.

Regarding claim 13, the prior art of record fail to disclose or suggest the claimed steps of: associating the document with a unique cluster identifier if the document is not a near-duplicate of any previously processed document and associating the document with a cluster identifier associated with the previously processed document if the document is a near-duplicate of a previously processed document in conjunction with the remaining, salient claim provisions.

Regarding claims 17, 26-29, 34, 35, 38, and 40, the prior art of record fail to disclose or suggest the claimed step of: generating at least two fingerprints in conjunction with the remaining, salient claim provisions.

Regarding claim 44 the prior art of record fail to discloses or suggest the claimed limitation of: documents in the collection of documents without any common fingerprints are not checked to determine whether or not they are near duplicates.

Claims 2-10, 12, 18-25, 36-37, 41-43, and 45 are dependent upon allowable base claims, therefore they are allowed in the conjunction with allowable claims.

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Kathrow U.S Patent No. 6,263,348 discloses method and apparatus for identifying the existence of differences between two files.

Burrows U.S Patent No. 5,745,900 discloses method for indexing duplicate database records using a full record fingerprint.

Aiken U.S Patent No. 6,240,409 discloses method and apparatus for detecting and summarizing document similarity within large document sets.

Levy U.S Patent No. 6,505,160 discloses connected audio and other media objects.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Merilyn P Nguyen whose telephone number is 703-305-5177. The examiner can normally be reached on M-F: 8:30 - 5:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Safet Metjahic can be reached on 703-308-1436. The fax phone numbers for the organization where this application or proceeding is assigned are 703-746-7239 for regular communications and 703-746-7240 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703-305-3900.

MN
March 6, 2003


SAFET METJAHIC
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100